Mr.Saurabh Thakare / International Journal of Engineering Research and Applications (IJERA)
ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 2, Mar-Apr 2012, pp.366-369

Intelligent Processing and Analysis of Image for Shot Boundary Detection

Mr.Saurabh Thakare
(Department of Electronics Engg. (Communication), P.I.E.T. Nagpur, Maharashtra (India)

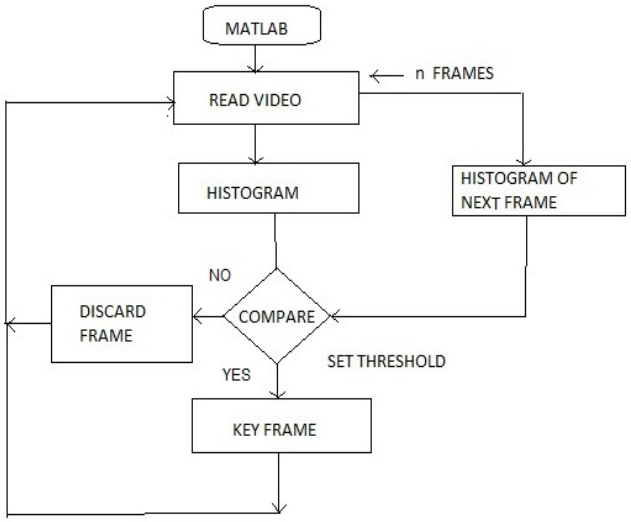
Abstract
Partitioning a video sequence into shot is the first step toward video-content analysis and content-based video browsing and retrieval. A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space. Shot boundary detection (SBD) plays an important role in content-based video retrieval. Shot is important and meaningful unit in videos. In this paper different video clips are used for further processing. Video consist of no. of frames depend on its size which represents content in video. As, wavelet transform have many feature over Fourier transform. Wavelet transform is use to compress the frames, it will help to minimise memory use to store the frames .Histogram technique is use to compare two frames. Key frame extraction is next part of paper .Key frame extraction is based on threshold calculation, mean deviation and standard deviation.

Keywords- Dynamic or static shot. , Key frame extraction, Shot boundary detection, Threshold, Wavelet transform

I. Introduction
Due to the rapid increase in the amount of video data generated and the wide range of video applications, an efficient and effective management of the data is necessary. Video shot boundary detection plays an important role in video processing. It is the first step toward video-content analysis and content-based video retrieval. A video shot consists of vary amount of frames. With the rapid development of computer and multimedia, video data surrounding people grows at an unprecedented speed. The rapidly expanding digital video information has motivated development of new technologies for efficient browsing, annotating and retrieval of video data. Content based video retrieval has attracted extensive research during the last decade. As the basis of video retrieval, shot boundary detection (SBD) is the first step of the video content analysis and content-based video retrieval .As video consist of no. of frames there is direct commands available for processing videos and mathematical analysis in digital image processing toolbox therefore one can easily implemented it, hence it is called intelligent processing. Wavelet transform is use to minimise memory requirement to store frames. Some researchers are making effort to find techniques that can improve the performance of shot boundary detection, such as using scale invariant feature transform (SIFT). SIFT was first presented by David G.Lowe in 2004 and has been well developed for object recognition and matching because its image features are invariant to image rotation, scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination. The video is segmented into shots, the most representative frames in each shot, named key frame, are extracted for further applications, such as video retrieval, user browsing and content analysis. So the extraction method of key frame in a shot is another key study direction. The easiest method is to take the first, middle or ending frame as key frame. This method results in one possibility that the extracted key frames have a low correlation in visual content. Another way is to split a shot into smaller video clips. Each clip can be a fixed time video or consecutive frame series characterized by high vision-content redundancy. Each clip typically includes several consecutive frames, but may also stretch to a complete shot. Then we can extract key frames from the clips in the way that the visual content redundancy is minimized. As the above mentioned analysis, this paper presents a unified framework for shot boundary detection and key frame extraction. In this paper, the key frame extraction is based on threshold calculation.

Different frames are extracted from video which are compressed using wavelet transform. Technical innovation by high-speed broadcasting and networking, digital video transmission, huge storage devices, etc., raises a demand for large-scale video database and its content-based access. These technologies are expected to enable "mining" knowledge from large-scale video archives that meets users' need. As a basic component of content-based video access, researchers tend to use feature-based similarity search for images and videos. However, several papers have recently been published which state that similarity search is getting more noisy and useless as the image video archives are getting larger, but instead, searching the "identical" image/ video is becoming useful . In this paper, histogram technique is for matching two blocks of frames.

II. Flow chart



"Fig". 1 Step's in algorithm

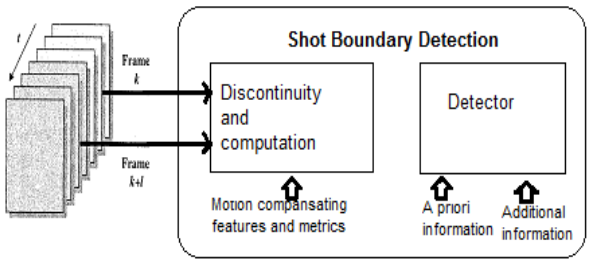
III. SHOT BOUNDARY DETECTION

III.I Image Segmentation

Each frame is divided into blocks with m rows and n columns. Then the difference of the corresponding blocks between two consecutive frames is computed. Finally, the final difference of two frames is obtained by adding up all the differences through different weights

III.II

Histogram plot is use to determine difference between blocks of two frame. Histogram is use to determine variation of pixel at different frequency of image.



"Fig".2 Shot Boundary Detection

IV. Wavelet transform

Wavelet compression is a form of data compression well suited for image compression. First a wavelet transform is applied. This produces as many coefficients as there are pixels in the image (i.e.: there is no compression yet since it is only a transform). These coefficients can then be compressed more easily because the information is statistically concentrated in just a few coefficients. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. A transform which localizes a function both in space and scaling and has some desirable properties compared to the Fourier transform. The transform is based on a wavelet matrix which can be computed more quickly than the analogous Fourier matrix.

V. Algorithm Description

Let F (k) be the k th frame in video, k =1, 2,.., n . The algorithm of shot boundary detection is described as follows.

Algorithm 1: Shot Boundary Detection

Step 1: Partitioning a frame into blocks with m rows and n columns, and B (i, j, k) stands for the block at (i, j) in the k th frame;

Step 2: Computing histogram matching difference between the corresponding blocks between consecutive frames in video sequence. H (i, j, k) and H (i, j, k+1) stand for the histogram of blocks at (*i, j*) in the kth and (k+1)th frame respectively. Block's difference is measured by the following equation

$$D_B(k,k+1,i,j)=\sum_{i=0}^{L-1}\frac{[H(i,j,k)-H(i,j,k+1)]^2}{H(i,j,k)}$$

..equation (1)

Where,

DB =Block difference,
L= Gray in image

Step 3: Computing histogram difference between two consecutive frames:

$$D(k,k+1)=\sum_{i=1}^{m}\sum_{j=1}^{n} Wij\ D_B(k,k+1,i,j)$$

..equation (2)

Where Wij stands for the weight of block at (i, j);

Step 4: Computing threshold computing the mean and standard variance of histogram difference over the whole video sequence. Mean and standard variance are defined as follows

$$MD=\sum_{k=1}^{n-1} D(k,k+1)\Big/ n-1$$

..equation (3)

$$STD=\sqrt{\sum_{k=1}^{n-1}(D(k,k+1)-MD)^2\Big/ n-1}$$

..equation (4)

Step 5: Shot boundary detection
Let threshold T=MD + a x STD Shot candidate detection: if D (i, i+1) is greater than equal to, the i^{th} frame is the end frame of previous shot, and the $(i+1)^{th}$ frame is the end frame of next shot.
Reference Frame: it is the first frame of each shot; General Frames: all the frames except for reference frame; "Shot Dynamic Factor" max (i): the maximum histogram within shot i; Dynamic Shot and Static Shot: a shot will be declared as dynamic shot, if its max(i) is bigger than MD; otherwise it is static shot; Fc(k) : the k^{th} frame within the current shot, k=1, 2, 3…F_{CN} (k) (F_{CN} (k) is the total number of the current shot) .The algorithm of key frame extraction is described as follows

Algorithm 2：Key Frame Extraction
Step 1: Computing the difference between all the general frames and reference frame with the above algorithm

$$Dc(1,k)=\sum_{i=1}^{m}\sum_{j=1}^{n} Wij\ DcB(1,k,i,j)...k=2,3,4,....FcN$$

..equation (5)

Step 2: Searching for the maximum difference within a shot:

$$max(i)=Dc(1,k)max,k=2,3,4....FcN$$

..equation (6)

Step 3: Determining "Shot Type" according to the relationship between max (i) and MD: Static Shot (0) OrDynamicShot

Shot Type ={1 if max(i) is greater than equal to MD}
{0 others}

..equation (7)

Step 4: Determining the position of key frame: if Shot Type=0, with respect to the odd number of a shot's frames, the frame in the middle of shot is chose as key frame; in the case of the even number, any one frame between the two frames in the middle of shot can be chose as key frame. If Shot Type=1, the frame with the maximum difference is declared as key frame.

VI.Experiment analysis

This paper gives idea how the key frame extracted .As video consist of number of frames depend upon size of video. These frames occupies large space in memory .So in this paper wavelet transform is used to compress frame size .Each frame is divided into sub-blocks .No. of rows and column selected in such way that it explains all the content in video. Frame rate is about 20 frames/sec. In the histogram methods, it is possible that the histograms of two frames are similar, but the contents are completely different. According to human beings' visual characteristic, if the difference of the same pixel of two frames is in a small range, we may think that these two pixels are the same; if the difference is larger than a threshold, we consider that they are different.

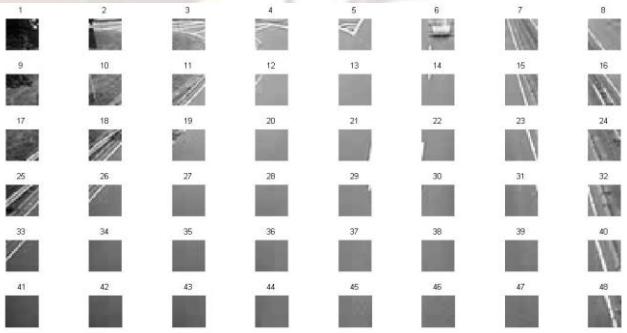
VII. Conclusion

In this paper extracted frames from the video compressed by using wavelet transform. Key frame extraction is depending upon threshold value calculated by using method discussed above. Key frames can satisfactorily represent the content of video. The wavelet transform is use to compress frames where video size is large. Each frame is divided into rows and column so that it represents all the content in frame and easy for next calculation part. In order to further improve accuracy, we will use multimodal information to segment video and generate video abstract. In this paper one important point is static or dynamic shot. Static shot referred to the picture or image which is not in motion. Dynamic is for continuously moving image.Comparision of these two shot help us to find out moving images in frames or steady images.
Multimodality-based video indexing will be our future direction.

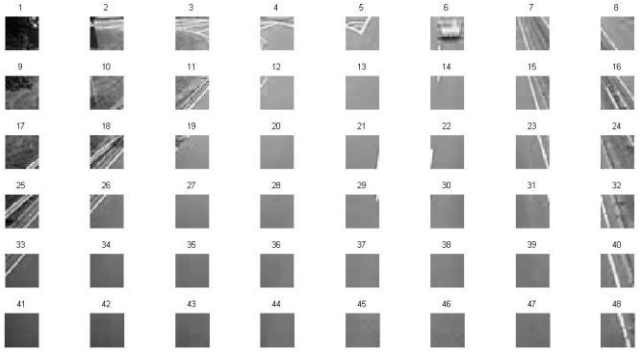
VIII.Result

The experiment result shows that key frames represent video content. The key frames are use to watch highlights of video. We use wavelet transform in this paper which are having many advantages over other image compression transform. So wavelet transform is use to compress the frame size, and it will helpful in application where video size is large. Also it saves the memory space required to store the frames. Dynamic and static shot is depending upon block difference and mean deviation value.Matlab software is use for completing and implementing algorithm with version 7.9.Consider two frames of video shown below.


Frame. 1 **Frame.2**



Frame .1 Divided into sub- blocks



Frame .2 Divided into sub –blocks

Frame 1. Block Difference: 2.087800e+005
Frame 2. Block Difference: 2.896984e+005

Threshold value is calculated using mean deviation and standard deviation formula discussed above and is compared with block difference to extract key frames.

ACKNOWLEDGEMENTS

The author would like to thank firstly, our GOD, and all friends who gave us any help related to this work. Finally, the most thanks to our family and our country INDIA.

REFERENCES

Journal Papers:

[1] N. Babaguchi, Y. Kawai,T. Ogura, and T. Kitahashi, *"Personalized abstraction of broadcasted American football video by highlight selection", IEEE Transaction On Multimedia, Vol.6,No.4,* August 2004, pp. 575-586.

[2] L. Zhu, M.S Guido, and Aggelos K.K., "*MIMAX optimal video summarization", IEEE Transaction on Circuits and System for Video Technology, Vol.15, No.10,* October 2005, pp. 1245—1256.

[3] A. Hanjalic, *"Shot-boundary detection: Unraveled and resolved?", IEEE Transaction on Circuits and System for Video Technology., Vol.12, No.2,* February, 2002, pp. 90-105.

[4] Z. Cernekova, I. Pitas, and C Nikou, *"Information theory-based shot cut/fade detection and video summarization", IEEE Transaction on Circuits and System for Video Technology, Vol.16, No.1,* January 2006, pp. 82-91.

Books:

[5] Rafael C. Gonzalez, *Digital Image Processing* (Low Price Edition –second edition Pearson Education Asia, 2002).

Proceedings Papers:

[6] A.Hanjalic, *"Multimodal approach to measuring excitement in video"*, Proceedings of International Conference on Multimedia and Expo ICME 03[C]. Vol.2, July 2003, pp. 289-292

[7] Y. Cheng, X. Yang, and D. Xu, *"A method for shot boundary detection with automatic threshold"*, TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communication, Control and Power Engineering[C], Vol.1, October 2002: 582-585.

[8] Y. Zhuang, Y. Rui, T.S. Huang, and S. Mehrotra, *"Adaptive key frame extraction using unsupervised clustering"*, Proceeding.ICIP'98[C], Chicago, IL,1998, Vol.1, pp. 866-870. vol. 9, pp. 3–19, Jan. 2000.